

Fourth-corner correlation is a score test statistic in a log-linear trait–environment model that is useful in permutation testing

Cajo J. F. ter Braak¹ 

Received: 6 May 2016 / Revised: 6 March 2017 / Published online: 27 March 2017
© The Author(s) 2017. This article is an open access publication

Abstract Ecologists wish to understand the role of traits of species in determining where each species occurs in the environment. For this, they wish to detect associations between species traits and environmental variables from three data tables, species count data from sites with associated environmental data and species trait data from data bases. These three tables leave a missing part, the fourth-corner. The fourth-corner correlations between quantitative traits and environmental variables, heuristically proposed 20 years ago, fill this corner. Generalized linear (mixed) models have been proposed more recently as a model-based alternative. This paper shows that the squared fourth-corner correlation times the total count is precisely the score test statistic for testing the linear-by-linear interaction in a Poisson log-linear model that also contains species and sites as main effects. For multiple traits and environmental variables, the score test statistic is proportional to the total inertia of a doubly constrained correspondence analysis. When the count data are over-dispersed compared to the Poisson or when there are other deviations from the model such as unobserved traits or environmental variables that interact with the observed ones, the score test statistic does not have the usual chi-square distribution. For these types of deviations, row- and column-based permutation methods (and their sequential combination) are proposed to control the type I error without undue loss of power (unless no deviation

Handling Editor: Pierre Dutilleul.

Electronic supplementary material The online version of this article (doi:[10.1007/s10651-017-0368-0](https://doi.org/10.1007/s10651-017-0368-0)) contains supplementary material, which is available to authorized users.

✉ Cajó J. F. ter Braak
cajo.terbraak@wur.nl

¹ Biometris, Wageningen University & Research, Wageningen, The Netherlands

is present), as illustrated in a small simulation study. The issues for valid statistical testing are illustrated using the well-known Dutch Dune Meadow data set.

Keywords Community ecology · Correspondence analysis · Fourth-corner · Permutation test · Score test statistic · Trait–environment association

1 Introduction

Ecological and evolutionary theory predicts that species have adapted to the environments they occupy (Southwood 1977; Townsend and Hildrew 1994). In recent years, understanding how the units of evolution (species) and their associated traits relate to the environment they inhabit has become a central focus in community ecology (McGill et al. 2006). A central question in this quest has been to establish the functionality of species traits, i.e. determine which traits allow species to survive and prosper where they do. Ultimately, by examining variation among attributes of species (traits) and among attributes of sites (environment), we can describe some of the important rules by which species assemblages emerge. As Legendre et al. (1997) stated: “Testing such hypotheses would require (1) a way to detecting associations between species and habitat characteristics, and (2) a way of testing the significance of these associations.” Traits and environmental (or habitat) variables cannot be correlated directly as they are measured on different units, namely species and sites, respectively, but can be connected via a non-negative link table with rows for sites and columns for species (e.g. presence–absence matrices, abundance or biomass information on species).

Legendre et al. (1997) developed a heuristic method referred to as the fourth-corner approach to trait–environment association. In that approach, three matrices containing information on the distributions of multiple species, species traits and the environmental attributes of species assemblages are combined to estimate a fourth matrix (the fourth-corner) containing correlations between traits and environment. Using ideas from multivariate analysis, Dolédec et al. (1996) developed a three-table ordination method, called RLQ, to establish the links between species traits and environmental variables. RLQ can consider either principal component analysis (PCA) or correspondence analysis (CA) as the central ordination method. For a single trait and a single environment variable, the version based on CA reduces to the fourth-corner method. The univariate version is by far the most used approach to link environmental and trait variation.

The three data tables at hand (Environment, Link and Trait data), denoted here by **E**, **Y** and **T**, respectively (Dolédec et al. 1996 used **R**, **L** and **Q**), can be arranged as

$$\begin{bmatrix} \mathbf{Y}^T & \mathbf{T} \\ \mathbf{E}^T & \mathbf{X} \end{bmatrix}$$

where **X** is the missing matrix (i.e. fourth-corner). With n sites, m species, p environmental variables and q traits, the dimensions of tables **Y**, **E** and **T** are $n \times m$, $n \times p$ and $m \times q$, respectively, so that **X** is $p \times q$. In the original method proposed by Legendre et al. (1997) the link table **Y** contained presence–absence of the m species in the n

sites but this was later generalized to abundance or count data by [Dray and Legendre \(2008\)](#). The link table is denoted here by \mathbf{Y} as it will be treated as response in a regression model later on.

The fourth-corner solution ([Dray and Legendre 2008](#); [Legendre et al. 1997](#)) is to determine \mathbf{X} in the simplest way, namely by the matrix product $\mathbf{X} = \mathbf{E}^T \mathbf{Y} \mathbf{T}$. For a nominal trait, a nominal environmental variable (expanded to indicator matrices \mathbf{E} and \mathbf{T}) and a presence–absence data table \mathbf{Y} , the fourth data table \mathbf{X} is simply a contingency table containing frequencies. A natural test statistic for significance testing is thus to compute the usual chi-square statistic, which will not necessarily follow a chi-square distribution because of the obvious dependencies between the entries. [Legendre et al. \(1997\)](#) investigated this issue and proposed permutation testing strategies as potential solutions. However, none of their strategies worked satisfactorily under all models they considered ([Dray and Legendre 2008](#)). Eventually, [ter Braak et al. \(2012\)](#) derived a strategy based on the sequential rejection principle ([Goeman and Solari 2010](#)) that controlled the type I error in data generated from any of models considered by [Dray et al. \(2014\)](#). This sequential strategy involves both row and column permutation (see Sect. 3).

For quantitative \mathbf{E} and \mathbf{T} , the same equation $\mathbf{X} = \mathbf{E}^T \mathbf{Y} \mathbf{T}$ can be used, except that the expansion to indicator matrices must now be replaced by normalization of each column of \mathbf{E} and \mathbf{T} to a weighted mean of zero and a weighed variance of 1, with site and species weights for \mathbf{E} and \mathbf{T} being the row and columns sums of \mathbf{Y} , respectively. This then yields a matrix \mathbf{X} consisting of fourth-corner correlations.

The motivation for the weighting came from considering an “inflated data table” ([Legendre et al. 1997](#)), in which \mathbf{Y} is vectorized, the zeroes removed, and each non-zero species–site combination is associated with the corresponding rows of \mathbf{E} and \mathbf{T} . The fourth-corner correlation between a trait and an environmental variable is then the Pearson correlation between the corresponding column of the trait in the inflated \mathbf{T} and the corresponding column of the environmental variable in the inflated \mathbf{E} . For more general non-negative data tables, this generalizes to a weighted Pearson correlation with abundances as weights and absences carrying zero weight ([Dray and Legendre 2008](#)).

The inflation process has an intuitive rationale when the abundance data are counts of individuals. The inflated data table simply lists all individuals (rows) and has, for $p = q = 1$, two variables (columns), namely the single trait and the single environmental variable. Each row has the trait value of the species it belongs to and the environmental value of the site which it inhabits. The fourth-corner correlation is then simply the unweighted Pearson correlation between the two variables of this table. This is a natural method to use when individuals are sampled with measurements of their traits and the environmental variables where they live. In such sampling, there will be intra-specific (and intra-site) variation, which is ignored in the original formulation of [Legendre et al. \(1997\)](#), but could certainly be accounted for. Significance testing of the correlation by permutation procedures proceeds similarly to the nominal variables case ([Dray et al. 2014](#)). The rationale followed by [Dolédéc et al. \(1996\)](#) to arrive at an equivalent solution is completely different; it is based on ways to constrain row and column scores in statistical triplets ([Cailliez and Pagès 1976](#); [Tenenhaus and](#)

Young 1985) defining a correspondence analysis. The link with (doubly) constrained correspondence analysis (Lavorel et al. 1998) returns at several places in this paper.

More recently, Pollock et al. (2012), Jamil et al. (2013) and Brown et al. (2014) independently proposed model-based approaches that generalize the fourth-corner problem to multiple traits and environmental variables using generalized linear (mixed) models (GL(M)M) for vectorized \mathbf{Y} , as in the bilinear regression approach of Gabriel (1998). Model-based approaches have great appeal (e.g. for nature conservation purposes) as they can improve the prediction of species abundances not only based on their environmental characteristics, but also on their traits and the interactions between traits and environment. These GL(M)M approaches allow for simultaneous modeling of the abundances of m species in terms of one or more traits and environmental variables. Mainstream methods of variable selection are used to build parsimonious models. In these approaches, \mathbf{X} becomes a matrix of (partial) regression coefficients estimating the direction and strength of the interaction between standardized traits and standardized environmental variables as well as their main effects on species distributions (Brown et al. 2014).

This paper seeks to establish connections between the earlier heuristic fourth-corner correlation and the more recent model-based approaches. One such, almost trivial, connection has been presented in the Appendix of Brown et al. (2014). In there, for a nominal trait and a nominal environmental variable, the fourth-corner \mathbf{X} is a contingency table obtained by merging columns and rows that belong to the same category of the trait and of the environmental variable, respectively, and the likelihood ratio test on interaction in a contingency table using a Poisson log-linear model is asymptotically equivalent with the usual chi-square test. No such relationships have been established for quantitative variables. The importance of such links is that they allow the generalization and unification of a simple and widely used heuristic method based on correlations (fourth-corner) to the GLM (fixed or mixed) regression machinery to link trait and environmental variation. This paper establishes that the squared fourth-corner correlation times the sum of the elements of the link table \mathbf{Y} (i.e. y_{++}) is precisely the score test statistic for testing the linear-by-linear interaction in a Poisson log-linear model with row and column main effects. Moreover, for multiple traits and environmental variables, the score test statistic is precisely y_{++} times the total inertia of a doubly constrained correspondence analysis (Kleyer 2012; Lavorel et al. 1999, 1998), which is the natural generalization of a singly constrained correspondence analysis, known as canonical correspondence analysis (Takane 2013; ter Braak 1986, 2014). It is also the natural generalization of RLQ (Dolédec et al. 1996; Dray et al. 2014) for correlated traits and environmental variables.

In ecological applications, however, the assumptions of the Poisson log-linear model are unlikely to hold true for a number of reasons. First, counts are typically over-dispersed compared to the Poisson and therefore modeled by, for example, a negative binomial distribution (Warton 2005). Second, observations from the same site are likely to be dependent and residual correlation among species is to be expected. This dependence has typically been addressed by resampling methods that resample entire sites instead of single individual observations (Oksanen et al. 2013; Wang et al. 2012). Third, observations on the same species are dependent when the observed environment interacts with unobserved (latent) traits, giving residual correlation among sites. This

dependence is accounted for in generalized linear mixed models for trait–environment interaction (Jamil et al. 2013; Pollock et al. 2012) by using a random slopes model. Warton et al. (2015a) extended such a random slopes logistic mixed model to a model with factor-analytic terms so as to account for both dependencies and analyzed it in the Bayesian framework using Gibbs sampling.

As this brief literature review shows, there are resampling-based (permutation and bootstrap) and model-based approaches to apply when model assumptions are unlikely to hold true. In the former the attempt is to overcome the shortcomings of the too simple model by resampling; in the latter, the simple model is extended until a ‘correct’ model has been found, defined as passing a number of diagnostics, so that one can then likely trust parametric (asymptotic) statistical inference. As an example, with model-based methods it is possible to build multi-trait multi-environment models in which the assumption of (conditional) independence is perhaps defensible. It is outside the scope of this paper to discuss the pros and cons of model-based versus resampling-based strategies, and how they might be combined. This paper takes the resampling approach using the simple Poisson model with interaction and shows by simulation that different deviations from the assumptions require different resampling methods to rescue the validity of the statistic test on trait–environment interaction. The different deviations also serve to explain why community-based and species-based inference (Shipley et al. 2007) may statistically yield different results (Ackerly et al. 2002; Peres-Neto et al. 2016) and why statistical tests based on community-based resampling as in Warton et al. (2015b) may have inflated type I error when the GLM-model does not hold true.

The paper is structured as follows. In Sect. 2 the score test statistic on interaction is derived, extended to the multi-trait multi-environmental variable case and also specialized to a number of common simple cases. In all cases, the score test statistic can be expressed in terms of the total inertia of a (doubly or singly) constrained correspondence analysis. In Sect. 3 the distribution of the test statistic is examined in the Poisson model from which it was derived and for five extended models and under four permutation schemes. Depending on the model, the permutation distribution obtained in a particular scheme does or does not correspond with the simulated distribution (i.e. the true distribution with sampling error) with only one scheme that controls the type I error in all models. This ‘max’ scheme, developed by ter Braak et al. (2012) from the sequential rejection principle, takes the maximum p value of the community-based permutation test and the species-based permutation test. Section 4 gives a real data example where, as in the simulations, community-based and species-based inference lead to different results, which can then be combined in the max scheme. Section 5 discusses the advantages, limitations and extensions of the approach taken in this paper and formulates the paradox that abundance is a weight in the fourth-corner correlation and a response in the log-linear model and that, nevertheless, these methods are closely related. The paradox is reconciled via a formula, well known in the literature on correspondence analysis, which expresses correspondence analysis as an approximation to a particular log-linear model and by noting that the fourth-correlation is the square-root of the only non-trivial eigenvalue of a doubly constrained correspondence analysis.

2 Theory

Unless otherwise noted, the response is assumed to be count data.

2.1 Likelihood and sufficient statistics

Arguably, the simplest statistical model used for detecting the trait–environment interaction is the log-linear model in which the count y_{ij} is assumed to follow a Poisson distribution with mean specified by

$$\log(\mu_{ij}) = r_i + c_j + b t_j e_i, \quad (1)$$

with r_i and c_j row (site) and column (species) main effects and b the coefficient measuring the direction and strength of the **t–e** interaction (i.e. the link between trait and environment). I derive the score test of the null hypothesis $b = 0$ with the alternative hypothesis $b \neq 0$. The row and column main effects are unknown nuisance parameters and as such they saturate the main effects for species and sites.

Under the assumption of Poisson distributed abundances $\{y_{ij}\}$, the relevant part of the log-likelihood is

$$l(\theta) = \sum_{i,j} \{y_{ij} \log(\mu_{ij}) - \mu_{ij}\} \quad (2)$$

with $E(y_{ij}) \equiv \mu_{ij} = \exp(r_i + c_j + b t_j e_i)$, so that

$$l(\theta) = \sum_i y_{i+} r_i + \sum_j y_{+j} c_j + b \sum_{i,j} y_{ij} t_j e_i - \mu_{++}, \quad (3)$$

where a “+” replacing an index means the sum over the index, e.g. $y_{i+} = \sum_j y_{ij}$. The minimal sufficient statistics are thus $\sum_{i,j} y_{ij} t_j e_i$ and the row and column totals $\{y_{i+}\}$ and $\{y_{+j}\}$.

2.2 Score test statistic

This subsection gives a recap of the score test (Bera and Bilias 2001; Cox and Hinkley 1974; Rao 1973; Yee 2015), which is simpler to compute than the likelihood ratio tests in most cases. The score function $U(\theta)$ is the derivative of the log-likelihood $l(\theta)$ with respect to the parameter vector θ :

$$U(\theta) = \frac{\partial}{\partial \theta} l(\theta), \quad (4)$$

which, under regularity conditions, is asymptotically normal with mean zero and variance equal to the Fisher information:

$$\text{var}_y[U(\theta)] = E_y \left[\frac{\partial}{\partial \theta} l(\theta) \frac{\partial}{\partial \theta^T} l(\theta) \right] = -E_y \left[\frac{\partial}{\partial \theta^T} U(\theta) \right] = I(\theta). \quad (5)$$

The score test statistic to test the null $H_0 : \theta = \theta_0$ versus $H_A : \theta \neq \theta_0$ is then

$$S(\theta_0) = U^T(\hat{\theta}_0) I^{-1}(\hat{\theta}_0) U(\hat{\theta}_0), \quad (6)$$

where $\hat{\theta}_0$ is θ_0 combined with the maximum likelihood estimate of the parameters that are not restricted in the null hypothesis $H_0 : \theta = \theta_0$.

2.3 Score test statistic for the interaction parameter

In the “Appendix” I derive the score test statistic for testing the trait–environment interaction, i.e. to test $H_0 : b = 0$ versus $H_A : b \neq 0$ in Eq. (1) for the case of the interaction between a single trait and single environmental variable. The score test statistic is

$$S(b)|_{b=0} = \frac{y_{++} \left[\sum_{i,j} y_{ij} \tilde{t}_j \tilde{e}_i \right]^2}{\sum_j y_{+j} \tilde{t}_j^2 \sum_i y_{i+} \tilde{e}_i^2} = y_{++} \text{cor}_Y^2(t, e) = y_{++} f^2 \quad (7)$$

with \tilde{t}_j and \tilde{e}_i centred versions of the trait and environmental variable:

$$\tilde{t}_j = t_j - \sum_j y_{+j} t_j / y_{++} \text{ and } \tilde{e}_i = e_i - \sum_i y_{i+} e_i / y_{++} \quad (8)$$

and f the fourth-corner correlation. This result is perhaps not unexpected because the squared fourth-corner correlation is the first eigenvalue of a doubly constrained correspondence analysis and it is known that (constrained) correspondence analysis decomposes the usual chi-square of a contingency table, χ^2 say, along factorial axes such that (Greenacre 2007, 1984; Takane 2013):

$$\chi^2 = y_{++} \sum_a \lambda_a = y_{++} (\text{total inertia}) \quad (9)$$

with λ_a the a th-eigenvalue and the sum is over all eigenvalues (constrained and unconstrained).

2.4 Score test for multiple traits and environmental variables

With data on p environmental variables and q traits in the $n \times p$ and $m \times q$ matrices $\mathbf{E} = \{e_{ik}\}$ and $\mathbf{T} = \{t_{jl}\}$, the interaction term in the log-linear model of Eq. (1) is replaced by a sum of $p \times q$ interaction terms $\{b_{kl}e_{ik}t_{jl}\}$. The score test statistic for the null hypothesis that all interaction coefficients are zero: $\mathbf{B} = \{b_{kl}\} = 0$, as derived in the “Appendix”, becomes

$$S(\mathbf{B})|_{\mathbf{B}=0} = y_{++} \text{tr}(\mathbf{D}^T \mathbf{D}) \quad (10)$$

with, after centring the columns of \mathbf{E} and \mathbf{T} as in Eq. (8),

$$\mathbf{D} = \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1/2} \mathbf{E}^T \mathbf{Y} \mathbf{T} \left(\mathbf{T}^T \mathbf{C} \mathbf{T} \right)^{-1/2}, \quad (11)$$

where \mathbf{R} and \mathbf{C} are diagonal matrices with the weights $\{y_{i+}\}$ and $\{y_{+j}\}$, respectively, on their diagonal.

This score test statistic, when divided by y_{++} , is equal to the total inertia of a doubly constrained correspondence analysis (Kleyer 2012; Lavorel et al. 1999, 1998), which is the natural generalization of a singly constrained correspondence analysis, known as canonical correspondence analysis (Takane 2013; ter Braak 1986, 2014). It is also the natural generalization of RLQ (Dolédéc et al. 1996; Dray et al. 2014) for trait and environmental data that are not \mathbf{R} - and \mathbf{C} -orthogonal.

The matrix \mathbf{D} can be re-expressed as well in terms of the residuals of \mathbf{Y} under the null model. The residual matrix is $\mathbf{Y}^* = y_{++} \mathbf{R}^{-1} \mathbf{Y} \mathbf{C}^{-1} - \mathbf{1}_n \mathbf{1}_m^T$ and, then,

$$\mathbf{D} = y_{++}^{-1} \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1/2} \mathbf{E}^T \mathbf{R} \mathbf{Y}^* \mathbf{C} \mathbf{T} \left(\mathbf{T}^T \mathbf{C} \mathbf{T} \right)^{-1/2} \quad (12)$$

which shows clearly the relation with a weighted regression of the residual matrix \mathbf{Y}^* on the traits and environmental variables where the weights are in \mathbf{R} and \mathbf{C} .

2.5 Special cases

There are important special cases of these results.

Trait and environment variables are factors or the identity matrix

If trait and environment variables are factors so that \mathbf{T} and \mathbf{E} are indicator matrices, the score test statistic is simply the usual chi-square statistic calculated from the contingency table \mathbf{Y}^{te} , say, containing the total abundance in each class of the cross-classification of the factor classes (see “Appendix”). This result was derived with saturated main effects (having free row and column parameters r_i and c_j). Brown et al. (2014) obtained a similar result from the log-linear model with \mathbf{T} and \mathbf{E} as main effects.

With \mathbf{T} and \mathbf{E} as diagonal matrices of size $m \times m$ and $n \times n$, respectively, the score test statistics becomes the usual chi-square statistic for contingency table \mathbf{Y} . This particular case is an analysis of \mathbf{Y} without external constraining information and has no value for trait–environment analysis.

With \mathbf{T} a diagonal matrix of size $m \times m$ and \mathbf{E} an $n \times p$ matrix, the score test statistic becomes y_{++} times the total inertia of a canonical correspondence analysis (ter Braak 1986). The (community-based) permutation test on the effects of environmental variables on species abundance using canonical correspondence analysis can thus be viewed as a test on the species-by-environment interaction in a log-linear model using a score test statistic.

Single trait and multiple environment variables

One of the first statistical methods to uncover and describe trait–environment association starts by calculating Community Weighted trait Means (CWM) (Kleyer 2012; Lavorel et al. 2008; Peres-Neto et al. 2016). With a single trait, the CWM is the single n -vector \mathbf{t}^* , with

$$\mathbf{t}^* = \mathbf{R}^{-1} \mathbf{Y} \mathbf{t}, \quad (13)$$

where, for simplicity of the following equations, \mathbf{t} is already \mathbf{C} -centred and \mathbf{C} -standardized as in equations (8) and (39) of the “Appendix”. The next step is then to calculate a regression of \mathbf{t}^* on the environmental variables. If a weighted regression is used, using weight matrix \mathbf{R} , the fitted values become

$$\hat{\mathbf{t}} = \mathbf{E} \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1} \mathbf{E}^T \mathbf{R} \mathbf{t}^* \quad (14)$$

and the regression sum of squares, more precisely, the weighted sum of squares of fitted values, is

$$\hat{\mathbf{t}}^T \mathbf{R} \hat{\mathbf{t}} = \mathbf{t}^{*T} \mathbf{R} \mathbf{E} \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1} \mathbf{E}^T \mathbf{R} \mathbf{t}^* = \mathbf{t}^T \mathbf{Y} \mathbf{E} \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1} \mathbf{E}^T \mathbf{Y} \mathbf{t} = \mathbf{D}^T \mathbf{D} \quad (15)$$

with \mathbf{D} as in Eq. (11). This shows that the score test statistic is y_{++} times the ratio of the regression sum of squares to the total sum of squares of \mathbf{t} . For a nominal environmental variable, this ratio is often called the squared correlation ratio, which is the statistic calculated in this situation by the function `fourthcorner2` in the R package `ade4` (Dray and Dufour 2007).

Note that the coefficient of determination of the weighted regression of \mathbf{t}^* on the environmental variables divides the regression sum of squares by the sum of squares based on \mathbf{t}^* (instead of on \mathbf{t}) and is thus a factor $\text{var}_R(\mathbf{t}) / \text{var}_R(\mathbf{t}^*)$ higher, as $\text{var}_R(\mathbf{t}^*) \leq \text{var}_R(\mathbf{t})$. This can give spuriously high coefficients of determination when there is in fact no relation at all. The reason is that $\text{var}_R(\mathbf{t}^*)$ is close to zero when there is no association between \mathbf{t} and \mathbf{Y} . Similarly, the simple and multiple correlation coefficients between \mathbf{t}^* and the environmental variable(s) are bad test statistics (Peres-Neto et al. 2016). The score test statistics derived in this paper do not have this shortcoming.

Multiple traits and single environment variable

The case of a single quantitative environment variable with multiple traits works analogously to the previous subsection with \mathbf{Y} transposed. In this case, weighted averages of the environmental variable are calculated for each species, resulting in an m -vector containing, what are called, species niche centroids. The vector can then be regressed on traits (Kleyer 2012; Šmilauer and Lepš 2014), analogously to the approach based on CWM.

3 Distribution of the score test statistic in permutation tests and extended models

This section examines by simulation the distribution of the score test statistic developed in the previous section when the assumptions of the Poisson log-linear model hold

true and shows that resampling methods that preserve the row and column totals of \mathbf{Y} yield a distribution of the score test statistic that is within sampling variation of both the asymptotic chi-square distribution and the simulated distribution. This section also shows that particular deviations from the assumptions of the Poisson model require different resampling methods to rescue the validity of the statistic test on trait–environment interaction. The deviations lead to models that serve to explain why community-based and species-based inference may statistically yield different results. In this paper the focus is on permutational methods of resampling.

The (asymptotic) distribution of the score test statistic is known to be chi-square with pq degrees of freedom (Cox and Hinkley 1974) when the statistical null model holds true, which is in our case the Poisson log-linear model (1) with $b = 0$. Appendix S1 provides code in the R language (R Core Team 2015) and results of simulations illustrating this. In these simulations, the analytical equations for the score test statistic using (7), (10), (15) and (43) give numerically the same value as the score test statistic calculated using the R package `mdscore` (da Silva-Junior et al. 2015); the difference between the likelihood ratio and the modified score statistic is small.

Figure 1 compares the true exceedance probability of the score test statistic as estimated on the basis of 10,000 simulated data sets (vertical axis) with the exceedance probability estimated by the chi-square distribution (parametric) and as obtained from four different permutation schemes using 999 permutations each (horizontal axis) across six different data-generating models (the 2×3 panels). Note that only the sixth model contains a true non-zero interaction between the observed trait \mathbf{t} and the observed environmental variable \mathbf{e} . Appendix S2 provides R-code for the simulations.

The first five models are different null models. The first model is the Poisson log-linear model (1) with $n = m = 30$, $b = 0$ and $r_i = \log(\mu_0) + \alpha e_i$ and $c_j = \beta t_j$, with $\mu_0 = 30$, $\alpha = 0.2$ and $\beta = 0.2$ (the main effects only model). In the second model, the distribution of the counts is set to negative binomial with variance function $\mu_{ij} + \mu_{ij}^2$. The next models deviate in one aspect from this second model (base model) by adding terms that induce correlations between species across sites and/or between sites across species. In the third model, the base model is extended with a random interaction term of the factor analytic form, $b_{zx}z_jx_i$, where $b_{zx} = 0.2$ and z_j and x_i are independent standard normal deviates. This model gives correlations both among species and among sites in any given data set. The next two models (in the second row of Fig. 1) also use independent standard normal deviates \mathbf{z} and \mathbf{x} (of length m and n) representing a latent trait and a latent environment variable, but now with interactions with the observed environment \mathbf{e} and the observed trait \mathbf{t} , respectively. In the fourth model, the base model is extended with the term $b_{ze}z_je_i$ ($b_{ze} = 0.2$), i.e. an interaction of a latent trait \mathbf{z} with the observed environment \mathbf{e} , which induces correlation among sites. This model is relevant when the observed environmental variable is known to influence the abundance of different species in different ways, but where these differences cannot be explained by the observed trait(s). The fourth model can be re-expressed as:

$$\log(\mu_{ij}) = \log(\mu_0) + \beta t_j + b_j e_i \quad (16)$$

with $b_j = \alpha + b_{ze} z_j$, that is, $b_j \sim N(\alpha, \sigma_b^2)$ with $\sigma_b^2 = b_{ze}^2$. As such, this model is a generalized linear mixed model with random species-specific slopes (random slopes model) as in [Jamil et al. \(2013\)](#). In the fifth model, the base model is extended with the term $b_{tx} t_j x_i$ ($b_{tx} = 0.2$), i.e. an interaction of a latent environmental variable \mathbf{x} with the observed trait \mathbf{t} , which induces correlation among species. This model is of interest when the observed trait is known to influence the abundance of different species in different ways, but where the differences cannot be explained by the observed environmental variables.

The sixth model is the only non-null (alternative) model, namely the base model extended with the terms $b_{te} t_j e_i + b_{ze} z_j e_i$ with $b_{te} = 0.2$.

The permutation tests ([Manly 2006](#)) are carried out using four different schemes (abbreviation between brackets) which all preserve the row and column totals of \mathbf{Y} .

1. (rc). Randomly permute both all rows and all columns of \mathbf{Y} in respect to each other ([Dolédéc et al. 1996](#)). This scheme was first proposed by [Welch \(1990\)](#) for permutation testing of interaction in balanced fixed-effects two-way analysis of variance. They destroy any relationship between \mathbf{Y} and \mathbf{E} , and \mathbf{Y} and \mathbf{T} , respectively.
2. (row). Randomly permute only the rows of \mathbf{Y} . This is model 2 of [Dray and Legendre \(2008\)](#), destroying any relationship between \mathbf{Y} and \mathbf{E} only.
3. (col). Randomly permute only the columns of \mathbf{Y} . This is model 4 of [Dray and Legendre \(2008\)](#), destroying any relationship between \mathbf{Y} and \mathbf{T} only.
4. (max). Perform a sequential test ([Goeman and Solari 2010](#)) with first the row-permutation test using scheme 2, and, if this test is significant, then the column-permutation test using scheme 3, or vice versa ([ter Braak et al. 2012](#)). In our case, both tests use the same score test statistic, Eq. (7), so that the sequential test (when both tests are carried out) is then equivalent with the test in which the final p value is the maximum of the two p values. Scheme 4 improves model 5 of [Dray and Legendre \(2008\)](#) and [Peres-Neto et al. \(2012\)](#) in the way the final p value is calculated.

On the basis of a suggestion of a reviewer, two more permutation methods, which permute the trait values (or the values of the environmental variable) in inflated tables, have been evaluated in Appendix S3.

Except in the bottom-right panel in Fig. 1 (i.e. non-null (alternative) model where $b_{te} \neq 0$), the ideal test in terms of type I error rates follows the 1:1 line. Lines above this line indicate liberal tests that have elevated Type I error rate (too many rejections at a specified nominal level, e.g. the horizontal dashed line at 0.05) and lines below the 1:1 line indicate conservative tests that have too few rejections at a specified nominal level. A test is said to control the type I error, if its type I error rate is at most the nominal level ([Goeman and Solari 2010](#)), that is, if its lines in Fig. 1 are all at or below the 1:1 line.

The exceedance probability based on the chi-square distribution with 1 degree of freedom is at the 1:1 line only for the Poisson model and is far above this line for the other models. In the top-row panels of Fig. 1, the rc, row and col schemes closely follow the 1:1 line, but the max scheme is slightly below this line, and thus conservative with an observed rejection rate in the 10,000 simulations of about 3% at the nominal

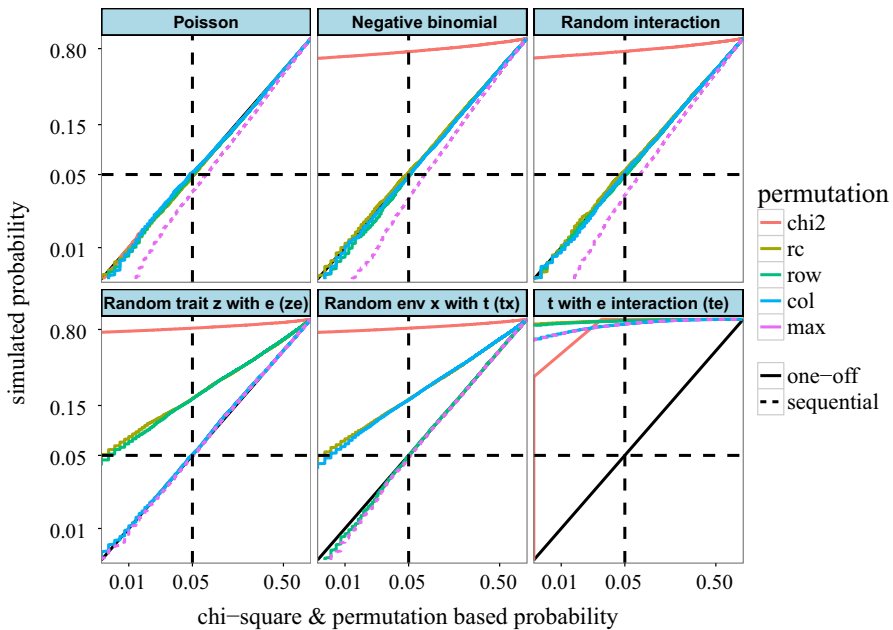


Fig. 1 Exceedance probability of the score test statistic of interaction in a Poisson log-linear model that also contains row and column main effects, as estimated from 10,000 simulated data sets, against the exceedance probability, as obtained from the chi-square distribution (chi2) and four different Monte Carlo permutation tests (999 permutations; permutation of rows and column simultaneously (rc), of rows (row) and of columns (col) and the sequential combination of the row and col scheme in the max scheme, which takes the maximum of the p values of the row and the col scheme). The panels show results for six data generating models, five of which represent typical deviations from the Poisson model (*top left*). All other models had negative binomially distributed response, without (*top middle*) and with further deviations (see text for details). The *bottom right* panel is the only model with a genuine non-zero interaction between the observed trait t and environment e . The variables z and x represent an unobserved (i.e. latent) trait and environmental variable, respectively

5% level of the test. Note, however, that the test is still reliable in the sense that it does not reject the null hypothesis more often than the nominal level.

In the first two panels of the bottom-row in Fig. 1, the rc scheme nearly coincides with the row scheme and the column scheme, respectively. The schemes are above the 1:1 line and thus liberal with a rejection rate of about 17% at the nominal 5% level of the test. In these panels, the max scheme nearly coincides with the column scheme and the row scheme, respectively. These schemes are about at the 1:1 line and thus have a rejection rate of about 5% at the nominal 5% level of the test.

The data generating model in the bottom-right panel is the only one containing a true non-zero interaction between the observed trait t and observed environmental variable e . In this case, the ideal line is Γ -shaped, indicating a high rejection rate at each nominal level of the test. At the nominal level of 5%, the rejection rate is ~ 0.90 for the col and max schemes and ~ 0.97 for the row and rc schemes. If b_{ze} is decreased from 0.2 to 0, these rejection rates are all > 0.98 in this case, and the plot is Γ -shaped, also for the chi-square based probability.

The conclusion from these simulations is that the use of the chi-square based probability gives highly inflated type I errors if the Poisson model does not strictly hold true. From the investigated permutation schemes (including the two methods of Appendix S3), the max scheme is the only scheme that controls the type I error in the five investigated models with latent variables, while providing a strong statistical power when $b_{te} \neq 0$.

4 Real data example

Different permutation schemes can also lead to different results in real data. This is illustrated here with the Dune Meadow data set (Jongman et al. 1995) consisting of abundances of 28 plants in 20 sites with five environmental variables and, from Jamil et al. (2013), five plant traits. The abundance is on a semi-quantitative rank scale with integer numbers from 0 (absent) to 9 (present everywhere). For illustration purposes only, abundances are treated as counts in this example and, alternatively converted to presence/absence. Suppose for a moment that the only available environmental variable is moisture, which is the major axis of variation of this data (Jongman et al. 1995), and one wishes to know whether it interacts with the plant trait SLA (specific leaf area). Using the fourth-corner score test statistic, the p values for the abundance data (with the p values obtained for presence/absence in this section between parentheses) for the permutation schemes rc, row and col are 0.008 (0.006), 0.028 (0.024), 0.218 (0.185), respectively (using 999 permutations). The first two schemes thus provide evidence for an interaction, whereas the col scheme does not. The simulations in Fig. 1 indicate that one possible reason for such a difference between the row and col schemes is that the environmental variable (moisture) interacts with a latent trait, even if that variable is independent of the trait of interest (SLA). There is indeed another trait in the Dune Meadow data set, namely seed mass, that has almost zero correlation with SLA ($r = -0.047$) and that interacts with moisture [p values of 0.0001 (0.0001) and 0.012 (0.0185) for the row and col schemes, respectively]. The p value of 0.028 (0.024) in the row scheme for the testing the interaction between SLA and moisture is thus likely caused by the interaction between seed mass and moisture. There is thus no evidence in these data that SLA and moisture have a real interaction. This example illustrates that the evidence for a trait–environment interaction is weak unless both the row and col schemes result in low p values. This line of reasoning leads naturally to the max scheme; the formal argument hinges on the theory of sequential testing (Goeman and Solari 2010) as given in ter Braak et al. (2012).

5 Discussion

This paper shows that the fourth-corner correlation, heuristically developed by Legendre et al. (1997) for examining trait–environment associations, has a close relationship with the Poisson log-linear model with interactions, which has recently been proposed as a model for trait–environment relationships (Brown et al. 2014; Warton et al. 2015b). The squared fourth-corner correlation is proportional to the score test statistic for testing the linear-by-linear interaction in the Poisson log-linear model with row and

column main effects. This result gives a mathematical underpinning of a conjecture that [Peres-Neto et al. \(2016\)](#) examined by simulation, namely that the fourth-corner correlation focuses on the interaction of a Poisson log-linear model and is not sensitive to main effects. Moreover, a score test is asymptotically equivalent with the likelihood ratio test, but much quicker to compute as it does not require fitting of the alternative model. This applies particularly to the test based on the fourth-corner correlation in comparison with the test based on the Poisson deviance difference between the main effects only model and the main effects with interaction model. In our R implementation, the test using the fourth-corner correlation is 140 times quicker to compute than the GLM-based test. Note that computing time easily becomes an issue with resampling for statistical inference, particularly, in large data sets.

Ecological data are likely over-dispersed. Then there are two popular models, the quasi-Poisson model and the negative binomial model. The quasi-Poisson model, with its variance proportional to the mean, allows a quasi-likelihood approach that leads to the Poisson deviance to be minimized and thus to the same estimates as the Poisson model. In this case, the squared fourth-corner is safe to use in resampling-based (permutation or bootstrap) significance tests. That is not the case for the negative binomial model (with variance function $\mu_{ij} + \phi\mu_{ij}^2$ and scale parameter ϕ). Then, the minimal sufficient statistics are the full data, instead of the three statistics below Eq. (3), and the score test statistic differs from the one in the Poisson model. Resampling based on the squared fourth-corner or the Poisson likelihood ratio (LR) is therefore no longer optimal and power may be lost. In a small simulation study as in the sixth panel of Fig. 1 (100 data sets per scenario and 99 permutations per data set), the power of the row, col and max schemes based on the negative binomial LR was 0.96, 0.94 and 0.93, respectively. By comparison, the power of the fourth-corner test on the same data sets was estimated as 0.97, 0.88 and 0.88, respectively, confirming some loss of power compared to using the negative binomial LR. The negative binomial LR is costly computationally and potentially numerically unstable; for example, in our implementation using the R package *mvabund* ([Wang et al. 2012](#)), I tried to obtain results for 1000 simulations with 999 permutation, but failed due to crashes of R. Note that the negative binomial GLM requires resampling for statistical inference as the parametric version inference is not very trustworthy, even in simple balanced design experiments for small to moderate data set sizes ([Szöcs and Schäfer 2015](#)). It would be of interest to develop a score test in the context of the negative binomial distribution.

Statistical tests in this paper have used resampling, based on restricted permutation of the counts. The restrictions ensured that the row and column totals were preserved. Without restrictions, permuting residuals would have been required to preserve these totals. Moreover, unrestricted resampling would treat the data or residuals as if they were exchangeable, whereas this is unlikely due to unobserved variation between species and/or sites.

[Brown et al. \(2014\)](#) advocate community-based resampling as being design-based. However, ecologists typically search for trait–environment association in observational studies. Therefore there exists no real design-based inference; the values at the sites or for species are in no way randomized by design. But it may still be hypothesized that values of traits, values of environmental variables or residuals from models

are exchangeable. This viewpoint supports both community-based and species-based resampling, although not necessarily completely random resampling when there is spatial or temporal autocorrelation or phylogenetic correlation.

Three types of restricted permutations were used here. The rc scheme permuted both rows and columns in the same resample, whereas the row and col schemes permuted either rows or columns. The simulation results showed that

- the rc scheme is not able to control the type I error rate when there is additional unobserved random variation among sites or among species that interacts with either the observed environment or the observed trait (as in the terms $b_{ze}z_j e_i$ and $b_{tx}t_j x_i$ in the simulation models, respectively).
- the row scheme is not able to control the type I error rate when there is additional species-based random variation that interacts with the observed environment (as in the term $b_{ze}z_j e_i$). In this scenario, the species respond differentially to the environment, but the differential response cannot be explained by the measured trait [see Eq. (16)]. By contrast, with additional site-based random variation, the row scheme controls the type I error rate, even if it interacts with the observed trait (as in the term $b_{tx}t_j x_i$).
- vice versa, the col scheme is not able to control the type I error rate when there is additional site-based random variation that interacts with the observed trait (as in the term $b_{tx}t_j x_i$). In this scenario, the species respond differentially to the trait, but the differential response cannot be explained by the measured environment. By contrast, with additional species-based random variation, the col scheme controls the type I error rate, even if it interacts with the observed trait (as in the term $b_{ze}z_j e_i$).
- the max scheme, in which the row- and column-based tests are combined, controlled the type I error rate in scenarios with either type of random variation.

Whereas the max scheme is perhaps currently the best simple method to test species-environment association, it is not yet perfect. In particular: 1) its type I error rate is below the nominal level when neither of these random effects is present (see top row of Fig. 1), resulting in some loss of power, and 2) its type I error rate can still be above the nominal level when both types of variation occur simultaneously. For example, if the base model in Fig. 1 is extended with huge latent interactions, namely $b_{tx}t_j x_i + b_{ze}z_j e_i$, with $b_{tx} = b_{ze} = 1$ (instead of 0.2 as in Fig. 1), the estimated type I error rate for the max scheme is 8.2% at the nominal 5% level (and the row and col scheme both give a type I error rate of 15%). So far, no resampling method has been found to control fully the type I error rate in this scenario. In this case, both the trait and the environment structure the species-by-site interaction, but do not interact among one another, at least not on the log-linear scale. To detect (or guard against) this scenario, the only way to go is presumably model-based (and Bayesian) as in [Warton et al. \(2015a\)](#).

Our simulation confirmed the remark of [Brown et al. \(2014\)](#) that community-based resampling “enables valid inferences that are robust to correlation between species, even when such correlation has not been incorporated into the fitted model”: the simulation in Fig. 1 (middle panel in second row) had correlations among species due a latent environmental variable x that was uncorrelated with the observed variable

e. The row scheme gave a correct type I error rate, but the col scheme did not, as species were correlated. Reversely, when there are dependencies among sites due to a latent trait \mathbf{z} , row-based resampling gave an inflated type I error rate, but column-based resampling did not (left panel in second row of Fig. 1). When either one or the other situation could be present, the max scheme is a solution to valid inference. When both situations are likely present, the max scheme also shows moderate type I error rate inflation and some form of p value adjustment estimated via simulation might be a way out (to undo possible type I inflation noted in the previous paragraph) or, the other elaborate option, explicit modeling of the correlations in a GLMM model. Both options are outside the scope of this paper. Of course, for observational data, any estimated correlation or association does not imply causation.

A reviewer raised serious objections against any permutation method that is based on permuting species by arguing that: “species (columns) are not the sampling units, they are out of the control of the experimenter and are generally assumed to be correlated due to species interactions and missing predictors”. I add phylogenetic relationships to this (see below). Therefore “Resampling species makes no sense from a design perspective, irrespective of the presence or absence of species-by-environment interaction effects”. The danger of all of this is that a statistical test using species-based resampling may have inflated type I error rate (is too liberal). Let me put this into the context of the max test. If the species-based resampling test is not performed, the final p value is the one from site-based resampling. The p value of the species part of the max test is then effectively nil (under the true null hypothesis, the null hypothesis is always rejected), which corresponds to the maximum type I error rate inflation possible. One is thus better off by applying the species-based test than by not applying it, even in the case that the above mentioned danger of some type I error rate inflation is real.

Note that, as yet, no valid GLM-based statistical test of species-by-environment interaction has been proposed. For example, the site-based residual bootstrapping approach of [Warton et al. \(2015b\)](#) suffers from the same type I error rate inflation as the simple site-based permutation scheme in the scenario of Fig. 1 that includes the $b_{ze}z_{je_i}$ term ([ter Braak et al. 2016](#)). Also this inflation can be counteracted by adding species-based resampling as in the max approach ([ter Braak et al. 2016](#)). Note also that missing predictors (either as main effects or interactions) are no problem as long as they do not interact with the observed trait and the observed environment. An example hereof is the random interaction scenario in Fig. 1.

Completely random permutations of species and/or of sites were used in this paper. This needs further adaptation as sites may be structured in space (spatial autocorrelation) and time (temporal autocorrelation) and species form a phylogeny (phylogenetic autocorrelation) so that neither sites nor species are really completely independent or exchangeable units. The net effect will be that the effective number of units is actually smaller than the number observed in the data (i.e. loss of degrees of freedom through autocorrelation), likely generating a liberal test when random permutations are used. Possible alternatives for random permutations are restricted permutations ([Lapointe and Garland 2014](#)) or data simulation that keeps the original spatial or phylogenetic structure in data ([Wagner and Dray 2015](#)). In this kind of hypothesis testing, phylogeny is treated as a nuisance: a trait–environment association is only judged valid when the association contributes beyond contributions due to phylogenetic relatedness.

For prediction, such a strong requirement is not needed. Prediction of abundance of a new species is expected to be better (with and without taking its trait value into account) the closer it is in the phylogeny to the species present in the data set.

The score test statistic for the testing the slope parameter in a simple regression is the sample size multiplied by the squared Pearson correlation (Bera and Biliias 2001). This result aligns nicely with the fourth-corner correlation defined as the Pearson correlation on inflated trait and environment data, but does not help to understand the link with the Poisson log-linear model. For this, the link between the fourth-corner correlation and correspondence analysis is more helpful as indicated in Sect. 2.3 and in more detail in the next paragraph.

The fourth-corner correlation and the log-linear model appear to treat species abundance in completely different ways: in the former as a weight (as fourth-corner is a weighted Pearson correlation) and in the latter as a response variable. In the following, this paradox is reconciled by using the relationship between the fourth-corner correlation and correspondence analysis (see Sect. 2). Recall that, for $p = q = 1$, the fourth-corner correlation arises as a doubly-constrained correspondence analysis in the RLQ-approach in Dolédec et al. (1996). It is well known that correspondence analysis is related to the Goodman's (1979) RC-model, which is the model of Eq. (1) with \mathbf{t} and \mathbf{e} latent. Indeed, a first order Taylor expansion of Eq. (1) in terms of $bt_j e_i$ yields the reconstitution formula of correspondence analysis (Greenacre 1984):

$$\mu_{ij} = R_i^* C_j^* \exp(bt_j e_i) \approx R_i^* C_j^* (1 + bt_j e_i), \quad (17)$$

where $R_i^* = e^{r_i}$ and $C_j^* = e^{c_j}$. So, for small b , both models can be expected to be very similar. Goodman (1981) showed that their estimation equation are then also very similar. For standardized \mathbf{t} and \mathbf{e} , b is the square-root of the first eigenvalue of correspondence analysis. Equation (17) applies, of course, also to our case of observed \mathbf{t} and \mathbf{e} . The log-linear model with row and column main effects and a linear-by-linear interaction is thus similar to the doubly constrained correspondence analysis. ter Braak (1985, 1988) showed such similarity also for data that follow the ecological niche model. Such unimodal data are very far from row-column independence and have in correspondence analysis a first non-trivial eigenvalue close to 1. That theory gives motivation to develop ordination methods for multiple traits and environmental variables based on a decomposition of the total inertia given in equations (10) and (11), as in the software package Canoco 5.1 (ter Braak and Šmilauer 2012), with row- and column-based permutation tests (of residuals) for statistical inference. Such methods can be used as a quick-scan of trait–environment associations. The alternative, or rather, the complementary approach is to go the full Bayesian model-based approach with latent variables and factor analytic structure of which Warton et al. (2015a) provide a nice first implementation. Even such models may need resampling methods, as even the simplest models using the negative binomial already need resampling for valid statistical inference for small to moderate data set sizes (Szöcs and Schäfer 2015). Between these extremes, there is room for GLM- and GLMM-based approaches that use row- and column-based resampling schemes for valid statistical inference.

Acknowledgements I would like to thank Pedro Peres-Neto and Stéphane Dray for discussions and suggestions and two anonymous reviewers for comments that improved the text.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Derivations

Score test statistic for the interaction parameter b

In this appendix, I derive the score test statistic for testing the trait–environment interaction, i.e. to test $H_0 : b = 0$ versus $H_A : b \neq 0$ in Eq. (1) in the main text for the case of the interaction between a single trait and single environmental variable. For this, the first and second order derivatives of the log-likelihood in Eq. (3) with respect to the parameters are required.

The score functions, which are partial derivatives of the log-likelihood $l(\theta)$ with respect to each of the parameters in the vector $\theta = (r_1, \dots, r_n, c_1, \dots, c_m, b)^T$, are

$$U(r_i) = \frac{\partial l(\theta)}{\partial r_i} = y_{i+} - \mu_{i+}, \quad (18)$$

$$U(c_j) = \frac{\partial l(\theta)}{\partial c_j} = y_{+j} - \mu_{+j}, \quad (19)$$

$$U(b) = \frac{\partial l(\theta)}{\partial b} = \sum_{i,j} (y_{ij} - \mu_{ij}) t_j e_i, \quad (20)$$

on using, for example,

$$\frac{\partial \mu_{ij}}{\partial b} = t_j e_i \mu_{ij}. \quad (21)$$

Under the null, $b = 0$, the score functions $U(r_i)$ and $U(c_j)$ for the row and column parameters, respectively, are all zero. For the score test statistic (6), the score function of b evaluated at $b = 0$ is needed:

$$U(b)|_{b=0} = \sum_{i,j} (y_{ij} - \hat{\mu}_{ij}^0) t_j e_i, \quad (22)$$

with $\hat{\mu}_{ij}^0 = y_{i+}y_{+j}/y_{++}$, the fitted values under row-column independence, and also the diagonal element for b of the inverse of the Fisher information. For the latter, the full Fisher information matrix $I(\theta)$ is needed, which consists of negative second order derivatives. Under the null hypothesis, the diagonal element for b is

$$I(\hat{\theta}_0)_{b,b} = -\frac{\partial^2 l(\theta)}{\partial^2 b} = \sum_{i,j} \hat{\mu}_{ij}^0 t_j^2 e_i^2 \quad (23)$$

and the off-diagonal elements for b with r_i and for b with c_j are

$$-\frac{\partial^2 l(\theta)}{\partial r_i \partial b} = \sum_j \hat{\mu}_{ij}^0 t_j e_i = y_{i+} e_i \sum_j y_{+j} t_j / y_{++} = y_{i+} e_i t_0 \quad (24)$$

$$-\frac{\partial^2 l(\theta)}{\partial c_j \partial b} = \sum_j \hat{\mu}_{ij}^0 t_j e_i = y_{+j} t_j \sum_i y_{i+} e_i / y_{++} = y_{+j} t_j e_0, \quad (25)$$

respectively, where t_0 and e_0 are the weighted trait and environmental means, respectively,

$$t_0 = \sum_j y_{+j} t_j / y_{++} \text{ and } e_0 = \sum_i y_{i+} e_i / y_{++}. \quad (26)$$

By centering the trait and environmental variables, i.e. by using $t_j - t_0$ and $e_i - e_0$ instead of t_j and e_i , the cross-elements for (b, r_i) and (b, c_j) are zero so that the inverse of the diagonal element for b of $I^{-1}(\hat{\theta}_0)$ is simply the corresponding element of the Fisher information of the centred data, i.e.

$$\left[I^{-1}(\hat{\theta}_0)_{b,b} \right]^{-1} = I(\hat{\theta}_0)_{b,b} = \sum_{i,j} \hat{\mu}_{ij}^0 (t_j - t_0)^2 (e_i - e_0)^2, \quad (27)$$

so that the score test statistics for the centered data is

$$\begin{aligned} S(b) |_{b=0} &= U^T(\hat{\theta}_0) I^{-1}(\hat{\theta}_0) U(\hat{\theta}_0) \\ &= \left[\sum_{i,j} (y_{ij} - \hat{\mu}_{ij}^0) (t_j - t_0) (e_i - e_0) \right]^2 / \left[\sum_{i,j} \hat{\mu}_{ij}^0 (t_j - t_0)^2 (e_i - e_0)^2 \right]. \end{aligned} \quad (28)$$

Note that $\sum_{i,j} \hat{\mu}_{ij}^0 (t_j - t_0) (e_i - e_0) = 0$ so that the term between square brackets in the numerator can be simplified to

$$\sum_{i,j} (y_{ij} - \hat{\mu}_{ij}^0) (t_j - t_0) (e_i - e_0) = y_{++} \text{cov}_Y(t, e) \quad (29)$$

where $\text{cov}_Y(t, e)$ is the weighted variance

$$\text{cov}_Y(t, e) = \sum_{i,j} y_{ij} (t_j - t_0) (e_i - e_0) / y_{++}, \quad (30)$$

and the denominator can be simplified to

$$y_{++}^{-1} \sum_j y_{+j} (t_j - t_0)^2 \sum_i y_{i+} (e_i - e_0)^2 = y_{++} \text{var}_C(t) \text{var}_R(e) \quad (31)$$

where the variances are weighted with respect to the row and column totals, i.e.

$$\text{var}_C(t) = y_{++}^{-1} \sum_j y_{+j} (t_j - t_0)^2 \text{ and } \text{var}_R(e) = y_{++}^{-1} \sum_i y_{i+} (e_i - e_0)^2. \quad (32)$$

The final result is

$$S(b) |_{b=0} = \frac{y_{++} \text{cov}_Y^2(t, e)}{[\text{var}_C(t) \text{var}_R(e)]} = y_{++} \text{cor}_Y^2(t, e) = y_{++} f^2 \quad (33)$$

where f is the fourth-corner correlation

$$f = \frac{\text{cov}_Y(t, e)}{[\text{var}_C(t) \text{var}_R(e)]^{1/2}}. \quad (34)$$

Score test for multiple traits and environmental variables

In this section, the score test statistic is derived for the case in which multiple traits and multiple environmental variables interact in driving species abundance. With data on p environmental variables and q traits in the $n \times p$ and $m \times q$ matrices $\mathbf{E} = \{e_{ik}\}$ and $\mathbf{T} = \{t_{jl}\}$, the interaction term in the log-linear model of Eq. (1) is replaced by a sum of $p \times q$ interaction terms $\{b_{kl}e_{ik}t_{jl}\}$. The null hypothesis is that all interaction coefficients are zero: $\mathbf{B} = \{b_{kl}\} = 0$. The score function for b_{kl} is

$$U(b_{kl}) |_{\mathbf{B}=0} = \sum_{i,j} (y_{ij} - \hat{\mu}_{ij}^0) t_{jl} e_{ik}. \quad (35)$$

To simplify equations, it is assumed from now onwards that each of the traits and environmental variables is centered (weighted mean 0) as in Eq. (26). Under the null hypothesis, the equation can then be simplified to

$$U(b_{kl}) |_{\mathbf{B}=0} = \sum_{i,j} y_{ij} t_{jl} e_{ik}. \quad (36)$$

The diagonal elements of the Fisher information matrix for the interaction coefficients $\{b_{kl}\}$ are all of the form of Eq. (23). The new off-diagonal elements involving two interaction coefficients, b_{kl} and $b_{k'l'}$, say, are

$$-\frac{\partial^2 l(\theta)}{\partial b_{kl} \partial b_{k'l'}} = \sum_{i,j} \mu_{ij} t_{jl} t_{j'l'} e_{ik} e_{ik'}. \quad (37)$$

Under the null hypothesis, $\hat{\mu}_{ij}^0 = y_{i+}y_{+j}/y_{++}$ can be inserted for μ_{ij} , resulting in

$$I(\hat{\theta}_0)_{b_{kl}, b_{k'l'}} = y_{++}^{-1} \sum_j y_{+j} t_{jl} t_{j'l'} \sum_i y_{i+} e_{ik} e_{ik'}. \quad (38)$$

The off-diagonal elements of $\{b_{kl}\}$ with r_i and c_j are of the form of Eqs. (24) and (25), which are all zero for centered data, so that the analogue of Eq. (27) holds true for

the submatrix of $I(\hat{\theta}_0)$ involving all interaction coefficients. To understand what the resulting score test statistic is, it is convenient to orthonormalize the trait and environmental data matrix in a weighted metric so that the corresponding submatrix of $I(\hat{\theta}_0)$ is a constant diagonal matrix (Takane 2013). The orthonormalizing transformations are

$$\mathbf{E}^* = \mathbf{E} \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1/2} \text{ and } \mathbf{T}^* = \mathbf{T} \left(\mathbf{T}^T \mathbf{C} \mathbf{T} \right)^{-1/2}, \quad (39)$$

where \mathbf{R} and \mathbf{C} are diagonal matrices with the weights $\{y_{i+}\}$ and $\{y_{+j}\}$, respectively, on their diagonal. By consequence, $\mathbf{E}^{*T} \mathbf{R} \mathbf{1}_n = \mathbf{0}_p$, $\mathbf{E}^{*T} \mathbf{R} \mathbf{E}^* = \mathbf{I}_p$, $\mathbf{T}^{*T} \mathbf{C} \mathbf{1}_m = \mathbf{0}_q$ and $\mathbf{T}^{*T} \mathbf{C} \mathbf{T}^* = \mathbf{C}_q$ and the score test statistic becomes

$$S(\mathbf{B})|_{\mathbf{B}=\mathbf{0}} = y_{++} tr \left(\mathbf{D}^T \mathbf{D} \right) \quad (40)$$

with

$$\mathbf{D} = \mathbf{E}^{*T} \mathbf{Y} \mathbf{T}^* = \left(\mathbf{E}^T \mathbf{R} \mathbf{E} \right)^{-1/2} \mathbf{E}^T \mathbf{Y} \mathbf{T} \left(\mathbf{T}^T \mathbf{C} \mathbf{T} \right)^{-1/2}. \quad (41)$$

The score test statistic can be re-expressed in analogy with Eq. (33) as

$$S(\mathbf{B})|_{\mathbf{B}=\mathbf{0}} = y_{++} tr \left(cov_Y(\mathbf{t}, \mathbf{e}) cov_R^{-1}(\mathbf{e}) cov_Y^T(\mathbf{t}, \mathbf{e}) cov_C^{-1}(\mathbf{t}) \right), \quad (42)$$

where \mathbf{e} and \mathbf{t} are now multivariate and the covariances are matrices.

Special case: T and E factors

One special case of these results is when trait and environment are both factors.

If trait and environment variables are factors, \mathbf{T} and \mathbf{E} are indicator matrices, when not yet centered, so that $\mathbf{E}^T \mathbf{Y} \mathbf{T}$ is a contingency table \mathbf{Y}^{te} , say, containing the total abundance in each class of the cross-classification of the factor classes. On inspecting the matrix multiplication $\mathbf{E}^T \mathbf{M}_0 \mathbf{T}$, where $(\mathbf{M}_0)_{ij} = \hat{\mu}_{ij}^0$, the expected values under the null hypothesis for \mathbf{Y} , $\hat{\mu}_{ij}^0 = y_{i+} y_{+j} / y_{++}$, become the usual expected values under row-column independence for \mathbf{Y}^{te} , $\hat{\mu}_{ij}^{0te} = y_{i+}^{te} y_{+j}^{te} / y_{++}^{te}$. Also, $\mathbf{T}^T \mathbf{C} \mathbf{T}$ and $\mathbf{E}^T \mathbf{R} \mathbf{E}$ are diagonal matrices, \mathbf{C}_t and \mathbf{R}_e , say, with the total abundances within factor classes, $\{y_{+j}^{te}\}$ and $\{y_{i+}^{te}\}$, respectively, on the diagonal.

To obtain their score test statistic, the inverses of $(\mathbf{E} - \mathbf{1}_n \mathbf{e}_0^T)^T \mathbf{R} (\mathbf{E} - \mathbf{1}_n \mathbf{e}_0^T)$ are also needed and similarly for \mathbf{T} . By completing the brackets and applying the Matrix Inversion Lemma (Brookes 2011; Rao 1973), it can be shown that the inverse is $(\mathbf{E}^T \mathbf{R} \mathbf{E})^{-1} + \Delta_e = \mathbf{R}_e^{-1} + \Delta_e$ and similarly for \mathbf{T} , giving a matrix Δ_t , and that $\Delta_e^T \mathbf{Y}^* \Delta_t = \mathbf{0}$. The resulting score test statistic is Eq. (40) with

$$\mathbf{D} = \mathbf{R}_e^{-1/2} \left(\mathbf{E}^T (\mathbf{Y} - \mathbf{M}_0) \mathbf{T} \right) \mathbf{C}_t^{-1/2} = \mathbf{R}_e^{-1/2} (\mathbf{Y}^{te} - \mathbf{M}_0^{te}) \mathbf{T} \mathbf{C}_t^{-1/2}. \quad (43)$$

The score test statistic is thus simply the usual chi-square statistic calculated from the contingency table \mathbf{Y}^{te} .

References

- Ackerly D, Knight C, Weiss S, Barton K, Starmer K (2002) Leaf size, specific leaf area and microhabitat distribution of chaparral woody plants: contrasting patterns in species level and community level analyses. *Oecologia* 130:449–457. doi:[10.1007/s004420100805](https://doi.org/10.1007/s004420100805)
- Bera AK, Biliyas Y (2001) Rao's score, Neyman's $C(\alpha)$ and Silvey's LM tests: an essay on historical developments and some new results. *J Stat Plan Inference* 97:9–44. doi:[10.1016/S0378-3758\(00\)00343-8](https://doi.org/10.1016/S0378-3758(00)00343-8)
- Brookes M (2011) The matrix reference manual. <http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/identity.html#InvLemma> (online)
- Brown AM, Warton DI, Andrew NR, Binns M, Cassis G, Gibb H (2014) The fourth-corner solution—using predictive models to understand how species traits interact with the environment. *Methods Ecol Evol* 5:344–352. doi:[10.1111/2041-210x.12163](https://doi.org/10.1111/2041-210x.12163)
- Cailliez F, Pagès J-P (1976) Introduction à l'Analyse des Données. Société de Mathématiques Appliquées et de Sciences Humaines, Paris
- Cox DR, Hinkley DV (1974) Theoretical statistics. Chapman and Hall, London
- da Silva-Junior AHM, da Silva DN, Ferrari SLP (2015) mdscore: an R package to compute improved score tests in generalized linear models. *J Stat Softw* 61:1–16. doi:[10.18637/jss.v061.c02](https://doi.org/10.18637/jss.v061.c02)
- Dolédéc S, Chessel D, ter Braak CJF, Champely S (1996) Matching species traits to environmental variables: a new three-table ordination method. *Environ Ecol Stat* 3:143–166
- Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* 22:1–20. doi:[10.18637/jss.v022.i04](https://doi.org/10.18637/jss.v022.i04)
- Dray S, Legendre P (2008) Testing the species traits-environment relationships: the fourth-corner problem revisited. *Ecology* 89:3400–3412. doi:[10.1890/08-0349.1](https://doi.org/10.1890/08-0349.1)
- Dray S, Choler P, Dolédéc S, Peres-Neto PR, Thuiller W, Pavoine S, ter Braak CJF (2014) Combining the fourth-corner and the RLQ methods for assessing trait responses to environmental variation. *Ecology* 95:14–21. doi:[10.1890/13-0196.1](https://doi.org/10.1890/13-0196.1)
- Gabriel KR (1998) Generalised bilinear regression. *Biometrika* 85:689–700. doi:[10.1093/biomet/85.3.689](https://doi.org/10.1093/biomet/85.3.689)
- Goeman JJ, Solari A (2010) The sequential rejection principle of familywise error control. *Ann Stat* 38:3782–3810. doi:[10.1214/10-AOS829](https://doi.org/10.1214/10-AOS829)
- Goodman LA (1979) Simple models for the analysis of association in cross-classifications having ordered categories. *J Am Stat Assoc* 74:537–552. doi:[10.2307/2286971](https://doi.org/10.2307/2286971)
- Goodman LA (1981) Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J Am Stat Assoc* 76:320–334. doi:[10.2307/2287833](https://doi.org/10.2307/2287833)
- Greenacre MJ (1984) Theory and applications of correspondence analysis. Academic Press, London
- Greenacre M (2007) Correspondence analysis in practice, 2nd edn. Chapman and Hall/CRC, London
- Jamil T, Ozinga WA, Kleyer M, ter Braak CJF (2013) Selecting traits that explain species-environment relationships: a generalized linear mixed model approach. *J Veg Sci* 24:988–1000. doi:[10.1111/j.1654-1103.2012.12036.x](https://doi.org/10.1111/j.1654-1103.2012.12036.x)
- Jongman RHG, ter Braak CJF, van Tongeren OFR (1995) Data analysis in community and landscape ecology. Cambridge University Press, Cambridge. doi:[10.2277/0521475740](https://doi.org/10.2277/0521475740)
- Kleyer M et al (2012) Assessing species and community functional responses to environmental gradients: which multivariate methods? *J Veg Sci* 23:805–821. doi:[10.1111/j.1654-1103.2012.01402.x](https://doi.org/10.1111/j.1654-1103.2012.01402.x)
- Lapointe F-J, Garland T Jr (2014) A generalized permutation model for the analysis of cross-species data. *J Classif* 18:109–127. doi:[10.1007/s00357-001-0007-0](https://doi.org/10.1007/s00357-001-0007-0)
- Lavorel S, Touzard B, Lebreton J-D, Clément B (1998) Identifying functional groups for response to disturbance in an abandoned pasture. *Acta Oecol* 19:227–240. doi:[10.1016/S1146-609X\(98\)80027-1](https://doi.org/10.1016/S1146-609X(98)80027-1)
- Lavorel S, Rochette C, Lebreton J-D (1999) Functional groups for response to disturbance in mediterranean old fields. *Oikos* 84:480–498. doi:[10.2307/3546427](https://doi.org/10.2307/3546427)
- Lavorel S et al (2008) Assessing functional diversity in the field—methodology matters!. *Funct Ecol* 22:134–147. doi:[10.1111/j.1365-2435.2007.01339.x](https://doi.org/10.1111/j.1365-2435.2007.01339.x)
- Legendre P, Galzin RG, Harmelin-Vivien ML (1997) Relating behavior to habitat: solutions to the fourth-corner problem. *Ecology* 78:547–562. doi:[10.2307/2266029](https://doi.org/10.2307/2266029)

- Manly BFJ (2006) Randomization, bootstrap and Monte Carlo methods in biology, 3rd edn. Chapman and Hall, London
- McGill BJ, Enquist BJ, Weiher E, Westoby M (2006) Rebuilding community ecology from functional traits. *Trends Ecol Evol* 21:178–185. doi:[10.1016/j.tree.2006.02.002](https://doi.org/10.1016/j.tree.2006.02.002)
- Oksanen J et al (2013) vegan: community ecology package R package version 20-9. <http://www.cranr-project.org/>
- Peres-Neto PR, Leibold MA, Dray S (2012) Assessing the effects of spatial contingency and environmental filtering on metacommunity phylogenetics. *Ecology* 93:S14–S30. doi:[10.1890/11-0494.1](https://doi.org/10.1890/11-0494.1)
- Peres-Neto PR, Dray S, ter Braak CJF (2016) Linking trait variation to the environment: critical issues with community-weighted mean correlation resolved by the fourth-corner approach. *Ecography*. doi:[10.1111/ecog.02302](https://doi.org/10.1111/ecog.02302)
- Pollock LJ, Morris WK, Veski PA (2012) The role of functional traits in species distributions revealed through a hierarchical model. *Ecography* 35:716–725. doi:[10.1111/j.1600-0587.2011.07085.x](https://doi.org/10.1111/j.1600-0587.2011.07085.x)
- R Core Team (2015) R: a language and environment for statistical computing, version 3.0. R Foundation for Statistical Computing. <http://www.R-project.org>, Vienna, Austria
- Rao CR (1973) Linear statistical inference and its application, 2nd edn. Wiley, New York
- Shipley B, Vile D, Garnier É (2007) Response to comments on from plant traits to plant communities: a statistical mechanistic approach to biodiversity. *Science* 316:1425–1425. doi:[10.1126/science.1140372](https://doi.org/10.1126/science.1140372)
- Šmilauer P, Lepš J (2014) Multivariate analysis of ecological data using CANOCO 5, 2nd edn. Cambridge University Press, Cambridge
- Southwood TRE (1977) Habitat, the templet for ecological strategies? *J Anim Ecol* 46:337–365. doi:[10.2307/3817](https://doi.org/10.2307/3817)
- Szőcs E, Schäfer R (2015) Ecotoxicology is not normal. *Environ Sci Pollut Res*. doi:[10.1007/s11356-015-4579-3](https://doi.org/10.1007/s11356-015-4579-3)
- Takane Y (2013) Constrained principal component analysis and related techniques. Chapman and Hall/CRC, London
- Tenenhaus M, Young FW (1985) An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis and other methods for quantifying categorical multivariate data. *Psychometrika* 50:91–119. doi:[10.1007/bf02294151](https://doi.org/10.1007/bf02294151)
- ter Braak CJF (1985) Correspondence analysis of incidence and abundance data: properties in terms of a unimodal response model. *Biometrics* 41:859–873. doi:[10.2307/2530959](https://doi.org/10.2307/2530959)
- ter Braak CJF (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167–1179. doi:[10.2307/1938672](https://doi.org/10.2307/1938672)
- ter Braak CJF (1988) Partial canonical correspondence analysis. In: Bock HH (ed) Classification and related methods of data analysis. Elsevier Science Publishers B.V. (North-Holland), Amsterdam, pp 551–558. <http://edepot.wur.nl/241165>
- ter Braak CJF (2014) History of canonical correspondence analysis. In: Blasius J, Greenacre M (eds) Visualization and verbalization of Data. Chapman and Hall/CRC, London, pp 61–75
- ter Braak CJF, Šmilauer P (2012) Canoco reference manual and user's guide: software for ordination, version 5.0. Microcomputer Power, Ithaca
- ter Braak CJF, Cormont A, Dray S (2012) Improved testing of species traits-environment relationships in the fourth-corner problem. *Ecology* 93:1525–1526. doi:[10.1890/12-0126.1](https://doi.org/10.1890/12-0126.1)
- ter Braak CJF, Peres-Neto PR, Dray S (2016) A critical issue in model-based inference for studying trait-based community assembly and a solution. *PeerJ* 5:e2885. doi:[10.7717/peerj.2885](https://doi.org/10.7717/peerj.2885)
- Townsend CR, Hildrew AG (1994) Species traits in relation to a habitat templet for river systems. *Freshw Biol* 31:265–275. doi:[10.1111/j.1365-2427.1994.tb01740.x](https://doi.org/10.1111/j.1365-2427.1994.tb01740.x)
- Wagner HH, Dray S (2015) Generating spatially constrained null models for irregularly spaced data using Moran spectral randomization methods. *Methods Ecol Evol* 6:1169–1178. doi:[10.1111/2041-210X.12407](https://doi.org/10.1111/2041-210X.12407)
- Wang Y, Naumann U, Wright ST, Warton DI (2012) mvabund—an R package for model-based analysis of multivariate abundance data. *Methods Ecol Evol* 3:471–474. doi:[10.1111/j.2041-210X.2012.00190.x](https://doi.org/10.1111/j.2041-210X.2012.00190.x)
- Warton DI (2005) Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics* 16:275–289. doi:[10.1002/env.702](https://doi.org/10.1002/env.702)
- Warton DI, Blanchet FG, O'Hara RB, Ovaskainen O, Taskinen S, Walker SC, Hui FKC (2015) So many variables: joint modeling in community ecology. *Trends Ecol Evol* 30:766–779. doi:[10.1016/j.tree.2015.09.007](https://doi.org/10.1016/j.tree.2015.09.007)

- Warton DI, Shipley B, Hastie T (2015b) CATS regression—a model-based approach to studying trait-based community assembly. *Methods Ecol Evol* 6:389–398. doi:[10.1111/2041-210x.12280](https://doi.org/10.1111/2041-210x.12280)
- Welch WJ (1990) Construction of permutation tests. *J Am Stat Assoc* 85:693–698. doi:[10.2307/2290004](https://doi.org/10.2307/2290004)
- Yee TW (2015) Vector generalized linear and additive models with an implementation in R. Appendix A, Springer, New York

Cajo J. F. ter Braak is personal professor at Biometris, Wageningen University & Research, the senior author of the software package Canoco for visualization of multivariate data and of the adaptive MCMC sampler Differential Evolution Markov Chain (DE-MC aka DREAM).